

КЛАССИЧЕСКИЕ КРИТЕРИИ ПРОВЕРКИ ГИПОТЕЗ ОБ ОДНОРОДНОСТИ ДИСПЕРСИЙ ПРИ НАРУШЕНИИ СТАНДАРТНЫХ ПРЕДПОЛОЖЕНИЙ

Горбунова А.А., Лемешко Б.Ю., Лемешко С.Б.
НГТУ, Новосибирск
E-mail: else-alice@rambler.ru

Критерии проверки гипотез об однородности выборок – это часто используемые в различных приложениях задачи статистического анализа. При этом речь может идти о проверке гипотез об однородности законов распределения, соответствующих анализируемым выборкам, или об однородности математических ожиданий, или об однородности дисперсий. Естественно, что наиболее полные выводы могут быть получены в первом случае, однако исследователя могут в большей степени интересовать вопросы о возможных отклонениях в средних значениях выборок или о различии в характеристиках рассеяния результатов измерений.

Одним из основных предположений при построении классических критериев проверки однородности дисперсий является принадлежность наблюдаемых случайных величин (погрешностей измерений) нормальному закону. Поэтому применение классических критериев всегда сопряжено с вопросом, насколько корректны получаемые выводы в данной конкретной ситуации? При нарушении предположения о принадлежности анализируемых величин нормальному закону условные распределения статистик критериев при справедливости проверяемой гипотезы, как правило, сильно изменяются.

Настоящая работа продолжает исследования устойчивости критериев проверки гипотез о равенстве дисперсий [1, 2]. Сравняются классические критерии Бартлетта, Кокрена, Фишера, Хартли, Левене, рассматриваются непараметрические (ранговые) критерии Ансари-Бредли, Муда, Сижела-Тьюки [3]. Цель работы заключалась:

– в исследовании распределений статистик перечисленных критериев при законах распределения наблюдаемых случайных величин, отличающихся от нормального;

– в сравнительном анализе мощности критериев относительно конкретных конкурирующих гипотез;

– в реализации возможности применения классических критериев в условиях нарушения предположений о нормальности случайных величин.

Проверяемая гипотеза о постоянстве дисперсий m выборок имеет вид

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_m^2, \quad (1)$$

а конкурирующая с ней гипотеза

$$H_1 : \sigma_{i_1}^2 \neq \sigma_{i_2}^2, \quad (2)$$

где неравенство выполняется хотя бы для одной пары индексов i_1, i_2 .

При исследовании распределений статистик, построении для этих распределений процентных точек и оценке мощности критериев относительно раз-

личных конкурирующих гипотез использовалась методика статистического моделирования и развиваемое программное обеспечение. Объем моделируемых выборок исследуемых статистик составлял величину $N=10^6$. При таких N величина разности между истинным законом распределения статистики и смоделированным эмпирическим по модулю не превышает величины 10^{-3} .

Исследования распределений статистик проводились при различных наблюдаемых законах, в частности, в случае принадлежности моделируемых выборок семейству с плотностью

$$De \theta_0 = f(x; \theta_0, \theta_1, \theta_2) = \frac{\theta_0}{2\theta_1 \Gamma(1/\theta_0)} \exp\left(-\left(\frac{|x - \theta_2|}{\theta_1}\right)^{\theta_0}\right) \quad (3)$$

при различных значениях параметра формы θ_0 . Это семейство может быть хорошей моделью для законов распределения погрешностей различных измерительных систем. Распределение $De \theta_0$ включает в качестве частных случаев распределение Лапласа $\theta_0 = 1$ и нормальное $\theta_0 = 2$. Семейство (3) позволяет задавать различные симметричные законы распределения в той или иной мере отличающиеся от нормального: чем меньше значение параметра формы θ_0 тем «тяжелее» хвосты распределения $De \theta_0$.

При сравнительном анализе мощности критериев рассматривались конкурирующие гипотезы вида $H_1: \sigma_m = d\sigma_0$. То есть конкурирующей гипотезе соответствует ситуация, когда $m-1$ выборка принадлежат закону с некоторым $\sigma = \sigma_0$, в то время как одна из выборок, например, с номером m имеет некоторую отличную дисперсию. Проверяемой гипотезе соответствует ситуация $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_m^2 = \sigma_0^2$.

Исследование мощности критериев Бартлетта, Кокрена, Хартли, Фишера и Левене относительно конкурирующих гипотез вида $H_1: \sigma_2 = d\sigma_0, d \neq 1$ при числе выборок $m=2$ показало, что в случае принадлежности случайных величин нормальному закону критерии Бартлетта, Кокрена, Хартли и Фишера имеют одинаковую мощность. Критерий Левене заметно им уступает.

При законах распределения, отличных от нормального, например, в случае принадлежности двух анализируемых выборок семейству законов распределения (3) критерии Бартлетта, Кокрена, Хартли и Фишера остаются эквивалентными по мощности, а критерий Левене заметно им уступает. Однако в случае законов с более «тяжелыми хвостами» (например, в случае принадлежности выборок распределению Лапласа) критерий Левене имеет преимущество в мощности.

Критерии Бартлетта, Кокрена, Хартли и Левене могут применяться при числе выборок $m > 2$. В таких ситуациях мощность этих критериев оказывается различной. При $m > 2$ в случае выполнения предположений о нормальном законе данные критерии можно упорядочить по убыванию мощности следующим образом:

Кокрена \succ Бартлетта \succ Хартли \succ Левене.

Порядок предпочтения сохраняется и в случае нарушений предположений о нормальном законе. Исключение касается ситуаций, когда выборки принадлежат законам с более “тяжелыми хвостами” по сравнению с нормальным законом. Например, в случае принадлежности выборок закону Лапласа критерий Левене оказывается несколько мощнее трех других.

Результаты исследований мощности ранговых критериев показали заметное преимущество критерия Муда и практическую эквивалентность критериев Ансари-Бредли и Сижела-Тьюки. Естественно, что непараметрические критерии уступают по мощности критериям Бартлетта, Кокрена, Хартли и Фишера.

Распределения статистик непараметрических критериев не зависят от вида закона, если обе выборки принадлежат одной и той же генеральной совокупности. Но, если выборки принадлежат разным законам, то при справедливости проверяемой гипотезы H_0 о равенстве дисперсий распределения статистик непараметрических критериев изменяются: они зависят от вида этих законов.

Результаты исследований показали предпочтительность критерия Кокрена, который в 2-х выборочном варианте не уступает любому другому, а в многовыборочном оказывается наиболее мощным (за исключением законов с тяжелыми хвостами). В этой связи для критерия Кокрена методами статистического моделирования были построены таблицы верхних процентных точек (1%, 5%, 10%) в случае принадлежности наблюдаемых величин распределениям семейства (3) при значениях параметра формы $\theta_0 = 1, 2, 3, 4, 5$ и различных объемах выборок n .

Настоящие исследования выполнены при частичной поддержке РФФИ (проект № 09-01-00056-а), Федерального агентства по образованию в рамках АВЦП “Развитие научного потенциала высшей школы” и ФЦП “Научные и научно-педагогические кадры инновационной России”.

Литература

1. Лемешко Б.Ю., Миркин Е.П. Критерии Бартлетта и Кокрена в измерительных задачах при вероятностных законах, отличающихся от нормального // Измерительная техника. – 2004. – №10. – С. 10-16.
2. Лемешко Б.Ю., Пономаренко В.М. Исследование распределений статистик, используемых для проверки гипотез о равенстве дисперсий при законах ошибок, отличных от нормального // Научный вестник НГТУ. – 2006. – №2(23). – С. 21-33.
3. Кобзарь А.И. Прикладная математическая статистика: Для инженеров и научных работников / А.И. Кобзарь. – М.: ФИЗМАТЛИТ, 2006. – 816 с.